Transmission of Realistic Sensation : Development of a Virtual Dome

Michitaka Hirose Kensuke Yokoyama Shin'ichi Sato
Department of Mechano-Informatics, Faculty of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
E-mail: hirose@ihl.t.u-tokyo.ac.jp

1 Introduction

Technology for transmitting the realistic sensations associated with a remote place is becoming a topic of growing concern in many applications such as telerobotics and teleconferencing. For future implementation of new telecommunication systems, Virtual Reality technology is expected to play a significant role. For example, the Head Mounted Display (HMD), a typical display device in Virtual Reality systems, can provide a very wide field of view, which is essential in providing visual sensation through the changing of displayed images depending on head movement.

Figure 1 shows the basic configuration of a telecommunication system with realistic sensation consisting of the HMD and a camera head driven by a spatial sensor attached to the HMD. As long as the distance between the HMD and the camera head is small, the basic system works very well. However, in situations where the camera head has to be located in a remote place, the time delay between head movements and displayed images becomes greater; causing considerable problems. If the time delay exceeds a certain limit, the user will be unable to feel suitable realistic sensations and may even lose his sense of orientation. In the worst case, the user may even suffer from syndromes such as sea-sickness.

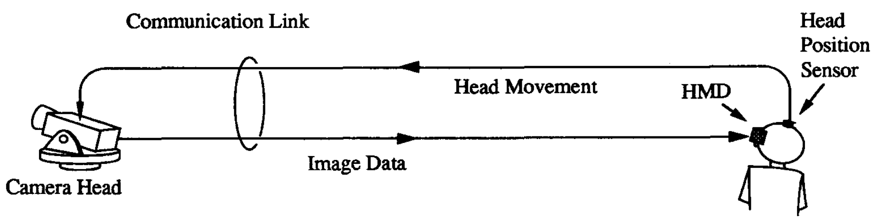


Figure 1: Concept of Conventional Telecommunication with HMD

2 Development of the Virtual Dome

2.1 Concept

To address the time delay problem, we proposed the design principle shown in Figure 2. The system was termed a "Virtual Dome". The key principle of the Virtual Dome was to separate the camera head movements and the user's head movements so that the time delay on the communication line did not directly affect the performance of the function associated with looking around using the HMD. By employing this method, the time delay effect could also be minimized.

The Virtual Dome includes 3 subsystems; (1) a camera head unit, (2) a communication line, and (3) an image display system. The camera head unit is located in the remote place

0-7803-1363-1/93 $3.00 © 1993 IEEE

and continually scans its surrounding space in order to capture a complete image of the surrounding area. Pictures are transmitted to the image display system via the communication line. The image display system consists of a graphics workstation and an HMD with a head position sensor. In the graphics workstation, a spherical object is constructed as a virtual screen and images taken by the camera head are texture-mapped onto it. By using the HMD, the user can experience remote synthetic sensations by looking around inside of this virtual omnimax screen. Since all of the information is located in the graphics workstation, there is no longer any time delay related to the looking around function. No matter how far away the camera head is located, the user will almost never lose the sensation of being in the remote place.

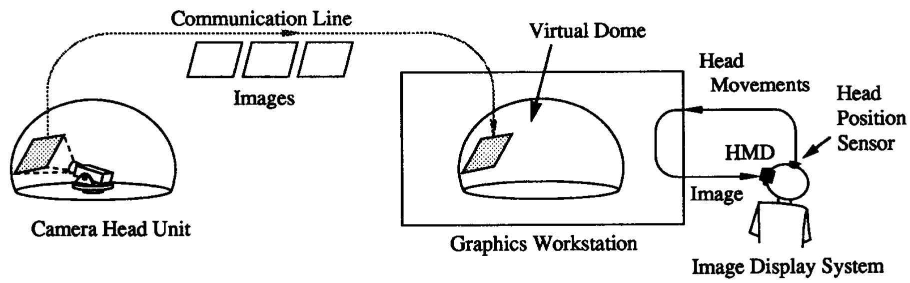


Figure 2: Concept of the "Virtual Dome"

2.2 Basic Implementation

The basic implementation is shown in Figure 3. The rotation of the camera head unit is driven by two stepping motors for horizontal and vertical movements in order to capture the surrounding information. Image capture is performed by stopping rotation in the continuous pattern shown in Figure 4. A CCD color camera (SONY XC-711RR) with 4 millisecond shutter speed is used to capture the images. The camera's viewing angle is 24 degrees in the horizontal direction and 18 degrees in the vertical direction. Under these conditions, the camera head unit can rotate at a rate of up to 6 rpm without blurring of the images.

The images are transferred to the graphics workstation (Silicon Graphics : IRIS 4D 210VGX) through the communication line, an NTSC channel in our current system. Next, a Frame Grabber (Silicon Graphics : Video Lab) is used to digitize these images. The image digitization presently being used has a resolution of 64×48 pixels and a depth of 24 bits per pixel (8 bits $\times$ 3 (RGB)). The reason for this relatively low resolution is that higher resolution requires more computation time and texture buffer space which consequently decreases the texture-mapping speed, lowering the performance of the graphics workstation.

Corresponding to the viewing angle of the camera, the dome has dimensions of 15 polygons in the horizontal direction and 5 polygons in the vertical direction (75 polygons in total). Hence, we can experience nearly every direction by looking around inside the dome. Approximate ranges are 360 degrees for the horizontal and 90 degrees for the vertical. The texture-mapped static image obtained through this system with the user wearing the HMD (VPL : Eyephone) is shown in Figure 5.

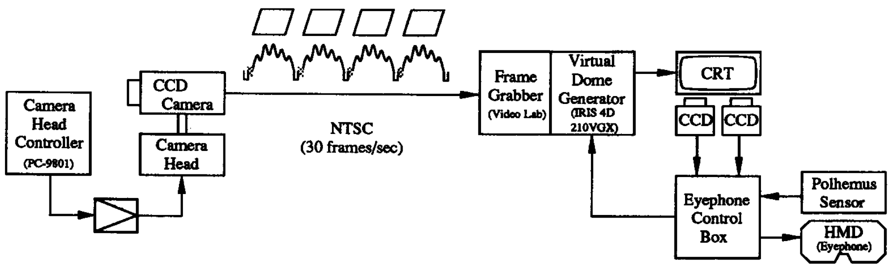


Figure 3: Basic Implementaion of the "Virtual Dome" System

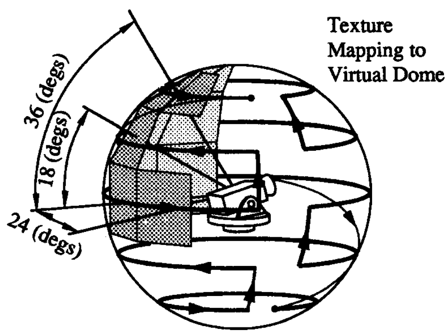


Figure 4: Shape of the Virtual Dome and Camera Movement

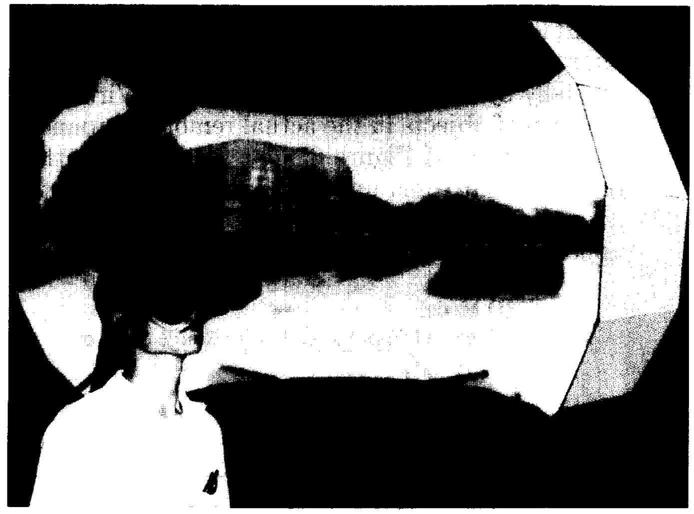


Figure 5 : Conceptual Image of the Virtual Dome

2.3 Refreshment of Images

For transmission of real-time information, further consideration is required due to the capacity limitations of the system. The limitations arise due to the time required to scan complete images of the surrounding space and then to draw the texture-mapped spherical dome in the graphics workstation. Therefore, a more efficient method for refreshing the images effectively would be to assign priority to refreshing images in the direction the user is currently looking.

Currently, a system has been implemented in which entire images of the surrounding space are first captured by the rotating camera head and sent to the workstation. The subsequent camera head actions are then driven by the head movements of the user with priority given to refreshing the images located in the user's scope of view. Hence, when the user changes his head orientation, he will initially see previously obtained images for a few instances. However, if the user continues looking in the same direction for a noticeable period of time, images from the remote camera will refresh the old ones and portray more recent images of the remote location. Using this approach, the user can essentially see live video images without losing his sense of orientation.

3 Extension to 3D Sensation

3.1 Concept

The next stage of Virtual Dome development was to add 3D cues, which like the wide field of view described above, are essential in generating realistic sensation. When using the HMD, the most important 3D cue is motion parallax. Presently, the image information obtained by rotating the camera head is basically 2D information. The user is only allowed to look around at his surroundings, in other words, there is no allowance for translational motion. A straightforward approach for adding a depth cue to 2D information is to translate the camera head according to the motion of the user. However, in this case, the time delay problem resurfaces again. Another approach would be to make the surface of the dome uneven. If approximate 3D information of the scene is obtained, realistic 3D sensation can then be partially synthesized by constructing a Virtual Dome with protrusions approximating the 3D form of objects in the actual remote surroundings as shown by the dotted line in the Top View diagram of Figure 6. The 2D images gathered by rotating the camera head are then projected onto this uneven Virtual Dome. In this case, from our experiences, even if the dome is only parially uneven, 3D sensation for the user is greatly increased with the images appearing proportional, even at extreme distances.

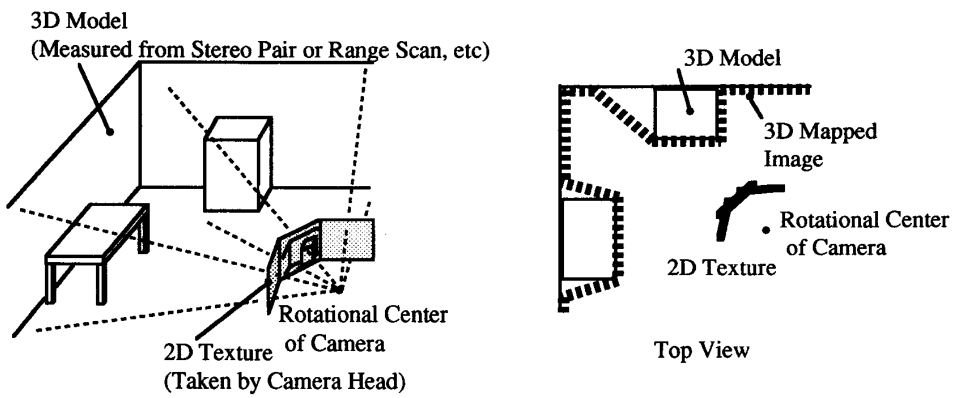


Fig 6: Generation of 3D Sensation from 2D Information and a 3D Model

3.2 Integration with Distance Information

Various computer vision software packages are possible candidates for carrying out the methods to calculate 3D information from 2D data. In our Virtual Dome system, the method which was originally developed by S. Tsuji et al. [3] was used. Figure 7 shows the principle behind this method. Suppose that (1) there is a marked point P in 3D space, (2) the camera rotates around a fixed axis, and (3) there is a barrier fixed in front of the camera head with 2 slits through which the surroundings can be seen. The barrier is fixed such that it rotates together with the camera head. Thus, when we rotate the camera head through an angle of 2θ in a counter-clockwise direction as viewed from above, the image of P projected through the left slit will disappear, and then reappear through the right slit of the barrier. Since θ is determined by the distance L, between P and the center of camera rotation, L can be obtained by using following formula:

$$L = \frac{R \sin \phi}{\sin(\phi - \theta)}$$

In our Virtual Dome system, the stepping motors with harmonic drive which drive the camera head unit were capable of rotating accurately at intervals of 0.0144 degrees. For our purposes, we rotated the camera head unit at intervals of 0.144 degrees. By rotating 2500 steps, images around the whole field of view could be obtained. Then, by arranging the one dimensional images obtained through the right and left slits, we could obtain a pair of panoramic images. The light intensity values obtained by scanning in one horizontal direction are shown in Figure 8. By using DP (dynamic programming) matching, it was possible to match angles by light intensity values. These paired angles theoretically corresponded to the same points in space. From the difference between the pairs of angles, the distance of the corresponding spatial point from the camera could be obtained. By repeating this operation for each scan line, 3D information could be obtained.

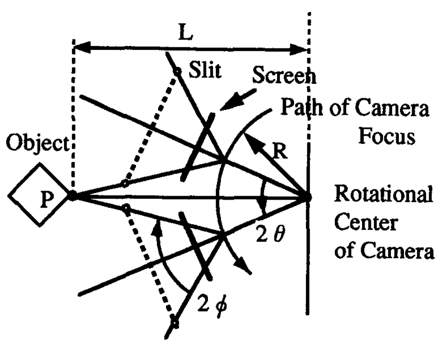


Figure 7 : Principle of the Slit Camera Method [3]

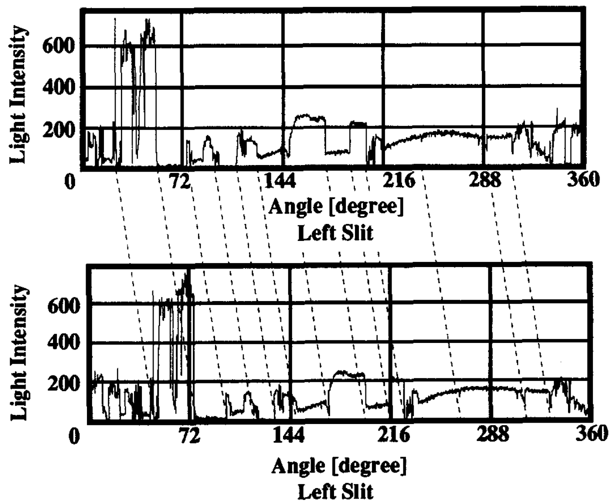


Figure 8 : Panoramic Images' Shading Value

By extracting edge points corresponding to sudden changes in light intensity of the panoramic image, and using only these edge points for DP matching, computing complexity could be greatly decreased and computing time significantly shortened. After obtaining the 3D information, 3D models approximating the surroundings could be generated. Figure 9 shows a typical result of 3D information measured by this method. The result shows that the method works well if the pair of angles corresponding to an edge are found properly and the point is near enough. Far-off distant information is not so important because motion parallax occurs only in the near vicinity of the user. At this time, only the points within 2 meters are used for generating the 3D model. The points which are further than 2 meters are projected onto the Virtual Dome surface in a normal fashion.

The next step was to use this 3D information to construct appropriate protrusions in the Virtual Dome. To simplify the problem, we calculated linearly hypothetical distances of points at regular angle increments based on the edge information obtained (left hand drawing of Figure 10). Thus, the distance information of the edge points was used only indirectly. However, by using this method, a Virtual Dome approximating the surroundings could be

simply constructed without inter-scanline search, and therefore the distortion of images projected onto this Virtual Dome was minimized. Figure 11 shows the 3D image generated by this method of using the two viewing directions (left and right) using a standardized angle increment of 2 degrees and a distance of 10 pixels between scanlines out of a total distance of 243 pixels (a total of 26 scanlines). Although not so neat, motion parallax synthesized from 2D images can be experienced. If we try to construct appropriate protrusions in the Virtual Dome with only edge information, failure to detect an edge point can cause considerable distortion of the 3D model. If (as shown in the right hand drawing of Figure 10) there is failure to detect some number of edge points on one scan line, the texture mapping process of matching edge points will end up matching points from different edges. A Virtual Dome constructed based on 3D information of this kind would have a very high degree of distortion.

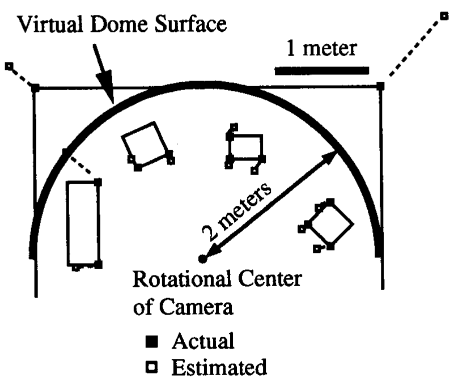


Figure 9 : Typical Result of 3D Measurement

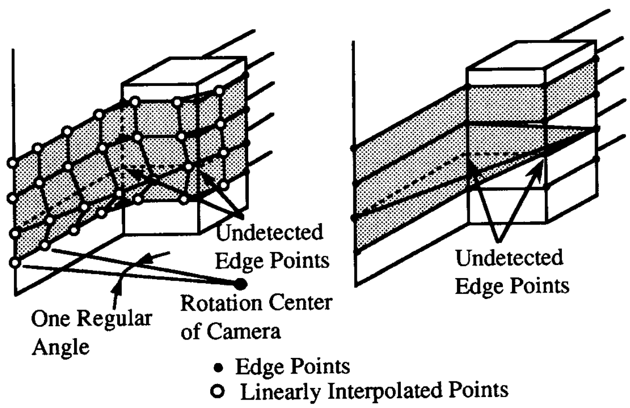
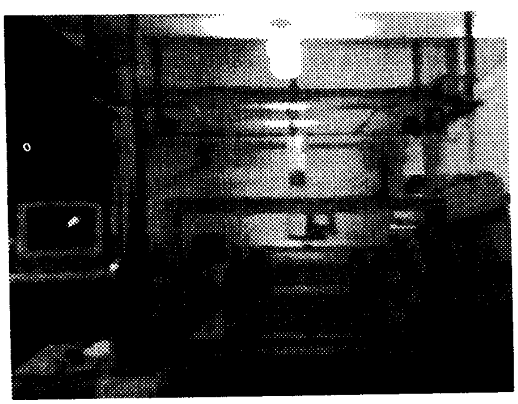


Figure 10 : Formation of 3D Model



View from Left Side

View from Right Side

Figure 11 : Motion Parallax Synthesized from 2D Images

4 Discussion

4.1 Problems of the Current System

The accuracy of the 3D information obtained by the stereo pair method is not so high for complex surroundings. If we apply this method to complex surroundings, distortion of the 3D

model becomes very large. Consequently, at the present stage, only 3D sensation of simple scenes can be transmitted. Furthermore, the process of obtaining the two panoramic images takes considerable time. Consequently, dynamic 3D information can not be transmitted by this method. The stereo pair method can only be used for static surroundings such as backgrounds.

4.2 Extension to Other Sensations

Sensations other than visualization are also very important. As a further extension of the Virtual Dome, we are also planning to add an auditory channel.

For example, a preliminary system for transmitting 3D auditory sensation is under development. Sound waves are analyzed by several microphone pairs attached to the camera head unit. From the phase shift between two sound waves detected by a microphone pair, the approximate direction of a sound source can be estimated. By using at least three microphone pairs, the 3D position of a sound source can be determined. In addition, if the loudness of the sound source is known, an even more precise estimation can be possible.

Once the source location is determined, the monaural sound and its location is transmitted to the user via the communication line. On the user's side, the 3D sound environment of a remote place can be re-generated by using a 3D sound system. In comparison with binaural recording, this system can significantly save channel capacity, and is very flexible because the information is highly coded.

In the same way as the auditory sensation transmission described here, if we have an appropriate recorder and player for other senses, wider ranges of sensation can be transmitted.

5 Conclusion

By using Virtual Reality technology, a communication system with realistic sensation can be implemented. This enables the user to see live images of a remote place without losing his sense of orientation. The most important element of this system is a buffering mechanism to overcome the time-delay problem which is harmful when trying to generate the realistic sensations of a remote place. Also, by adding 3D cues to the Virtual Dome (although not so accurate) 3D sensation can be sufficiently synthesized.

References

[1] R.Kijima and M.Hirose, "'Virtual Science' of Accuracy in Generated Environments — Focusing on the Effect of Time Delay in Virtual—", Human Interface News and Report, Vol.6, No.2, pp.140-145, 1991.

[2] M.Hirose, K.Hirota, R.Kijima, M.Kanno, K.Hayakawa and K.Yokoyama, "A Study on Synthetic Visual Sensation through Artificial Reality", Proceedings of the 7th Symposium on Human Interface, pp.675-682, 1991.

[3] H.Ishiguro, M.Yamamoto and S.Tsuji, "Omni-directional stereo for making global map", Porc. 3rd Int. Conf. Comput. Vision pp.540-547, 1990.

[4] Y.Ohta and T.Kanade, "Stereo by intra- and inter- scan line search using dynamic programming", IEEE Trans, PAMI-7,2, pp.139-154, 1985.